Neuron, Vol. 48, 139–148, October 6, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.neuron.2005.08.039

Perceptual Organization of Tone Sequences in the Auditory Cortex of Awake Macaques

Christophe Micheyl,[1,*] Biao Tian,[2] Robert P. Carlyon,[3] and Josef P. Rauschecker[2,*]

[1]Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

[2]Department of Physiology and Biophysics
Georgetown Institute for Cognitive and Computational Sciences
Georgetown University Medical Center
Washington, District of Columbia 20007

[3]MRC Cognition and Brain Sciences Unit
Cambridge CB2 2EF
England

Summary

Acoustic sequences such as speech and music are generally perceived as coherent auditory "streams," which can be individually attended to and followed over time. Although the psychophysical stimulus parameters governing this "auditory streaming" are well established, the brain mechanisms underlying the formation of auditory streams remain largely unknown. In particular, an essential feature of the phenomenon, which corresponds to the fact that the segregation of sounds into streams typically takes several seconds to build up, remains unexplained. Here, we show that this and other major features of auditory-stream formation measured in humans using alternating-tone sequences can be quantitatively accounted for based on single-unit responses recorded in the primary auditory cortex (A1) of awake rhesus monkeys listening to the same sound sequences.

Introduction

When we listen to a speaker in a crowd or follow a musical instrument in an orchestra, we rely unconsciously on our brain's ability to organize highly complex acoustic sequences into perceptual "streams" (Bregman, 1990; Bregman and Campbell, 1971; Carlyon, 2004). Streams are among the "objects" of audition. Like the objects of visual perception, they generally correspond to physical objects in the environment—in this case, sound sources—and as perceptual entities, they can be selectively attended to, processed, and followed over time (Bregman, 1990; Carlyon, 2004; Micheyl et al., 2005). Behavioral studies indicate that several animal species, including monkey (Izumi, 2002), bird (Hulse et al., 1997; MacDougall-Shackleton et al., 1998), and fish (Fay, 1998), experience this "auditory streaming" phenomenon. This supports the view that it represents a widespread and fundamental perceptual-organization ability, likely to be of crucial importance for survival in diverse ecological environments where multiple sound sources are often present and need to be parsed.

In the laboratory, the formation of auditory streams can be demonstrated simply by using repetitive sequences of tones alternating between two frequencies, A and B, as illustrated in Figure 1A. When the frequency separation (ΔF) between A and B is small and the tones alternate slowly, listeners hear a coherent sequence of tones whose pitch jumps up and down. In contrast, when ΔF is large and/or the repetition rate (RR) is fast, two parallel but separate streams of constant-pitch tones are heard.

While various theories and computational models have been proposed to explain auditory stream formation (Anstis and Saida, 1985; Beauvois and Meddis, 1991; Bregman, 1990; Kanwal et al., 2003; Hartmann and Johnson, 1991; McCabe and Denham, 1997; van Noorden, 1975), the actual brain mechanisms underlying this perceptual phenomenon remain unclear. Possible brain correlates of auditory streaming have been sought in earlier studies using auditory event-related potentials in humans (Alain et al., 1998; Hung et al., 2000; Jones et al., 1998; Näätänen et al., 2001; Sussman et al., 1999) or single-unit recordings in animals (Bee and Klump, 2004; Fishman et al., 2001, 2004; Kanwal et al., 2003). However, the conclusions of these earlier studies are limited by the fact that the neural response patterns putatively associated with the one- and two-stream percepts were always induced using physically different stimuli. Stimulus-induced changes in neural response patterns are a common confound in studies attempting to bridge the gap between neural activity and perception (Logothetis and Schall, 1989; Parker and Newsome, 1998).

Here, we take advantage of the fact that stream segregation generally takes some time to build up (Anstis and Saida, 1985; Bregman, 1978; Carlyon et al., 2001); initially, sound sequences tend to be heard as a single, fused stream, and it is only after several seconds that the sounds appear to split into distinct streams that can be individually followed. Thus, under appropriate conditions, the percept evoked by a physically unchanging sequence of alternating tones tends to switch from that of a single stream to that of two streams. This systematic change in auditory percept over time makes it possible to compare neural responses associated with dramatically different percepts without any change in the evoking stimulus, simply by recording neural responses at various points during an ongoing sequence of sounds. Another important feature of our approach, which distinguishes it from earlier work on the neural basis of auditory stream segregation, is that the statistical variability of the neural responses is used to predict percept probabilities. Using this approach, we demonstrate a striking correspondence between the temporal dynamics of neural responses to alternating-tone sequences in the primary auditory cortex (A1) of awake rhesus monkeys and the perceptual build-up of auditory stream segregation measured in humans listening to similar sound sequences.

*Correspondence: cmicheyl@mit.edu (C.M.); rauschej@georgetown.edu (J.P.R.)

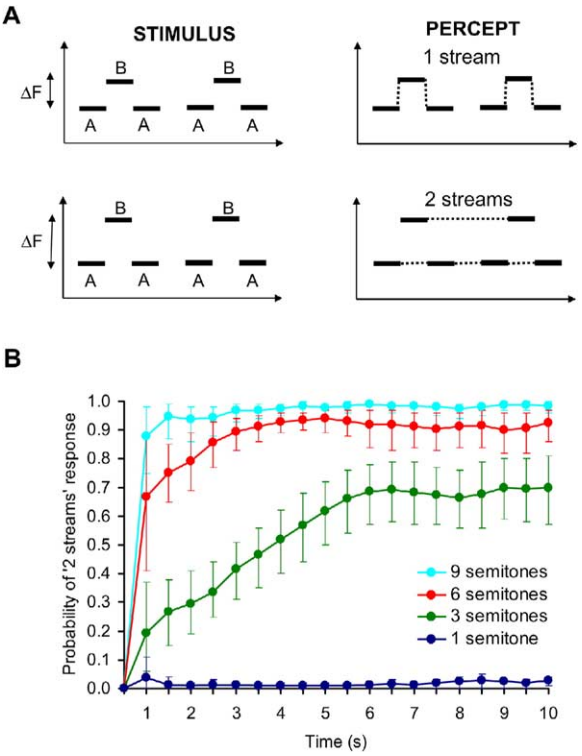

Figure 1. Auditory Streams: Stimuli and Percepts

(A) Stimulus sequences used to measure perceptual auditory streaming (left) and corresponding auditory percepts (right). The stimulus sequences consisted of repeating tone triplets, ABA, where A and B represent tones of different frequencies. The top panels illustrate the case of a small frequency separation (ΔF) between the A and B tones; the corresponding percept is that of a single stream of connected tones, with a distinctive galloping rhythm: ABA-ABA-…, where the dash stands for a silent gap. The lower panels illustrate the case of a large ΔF between the A and B tones; this typically evokes a percept of two monotonic streams with different tempi playing in parallel: A-A-A-A… on the one hand, -B—B on the other.

(B) The build-up of auditory stream segregation. The different curves represent the proportion of "two streams" responses measured psychophysically in human subjects listening to sound sequences like those illustrated in Figure 1A, for different frequency separations (ΔFs) between the A and B tones. The abscissa indicates time since sequence onset (total sequence duration: 10 s). The error bars indicate 95%-confidence intervals around the mean proportions estimated using statistical bootstrap.

Results

Psychophysical Demonstration of the Build-Up of Auditory Streaming

To demonstrate the build-up of stream segregation, we presented ten human listeners with 10 s sequences of repeating-tone triplets like those illustrated in Figure 1A. The listeners were instructed to indicate as promptly as possible after the onset of each sequence whether they heard one or two streams and to update their response each time the percept switched, until the end of the sequence. The responses were recorded along with their time of occurrence and averaged across stimulus presentations and listeners to produce psychometric functions representing the probability of hearing two streams as a function of time after sequence onset. This was done at four different ΔFs, chosen to yield different build-up rates and asymptotic probabilities of segregation. The "A" tone frequency varied between 500 and 4000 Hz across listeners. The resulting psychometric functions did not depend on the A tone frequency and were averaged across this parameter to form Figure 1B.

It can be seen that at the lowest ΔF tested (1 semitone) listeners very rarely reported hearing the sequence as segregated; throughout the entire sequence duration, the probability of a "two streams" response remained close to (and, in fact, not significantly different from) zero. In contrast, at the largest ΔF tested (9 semitones), listeners nearly always heard the sequence as segregated; the probability of a "two streams" response increased markedly within the first 2–3 s of stimulation, and it thereafter remained close to (and, in fact, not significantly different from) one. At the two intermediate ΔFs (3 and 6 semitones), judgments were more variable across listeners (as indicated by the larger confidence intervals) as well as less consistent within a given listener: the same sequence could be judged as "two streams" on one presentation but as "one stream" on another, resulting in intermediate asymptotic proportions of "two streams" responses. The time taken for the percept to switch from one to two streams was also more variable across sequence presentations and listeners at the 3 and 6 semitone ΔFs (across-listener SDs of 1.32 s and 1.23 s, respectively) than at the 9 semitone ΔF (SD = 0.46 s). Most listeners usually experienced a single switch per sequence, even at intermediate ΔFs; the average number of switches per sequence was 0.90 at the 3 semitone ΔF, 0.55 at the 6 semitone ΔF, 0.23 at the 9 semitone ΔF, and 0.08 at the 1 semitone ΔF. A possible reason for these small numbers of reversals is that the sequence duration used here was not much longer than the time required for the probability of segregation to build up in the intermediate ΔF condition (3 semitones), at which the percept was potentially the most unstable. Despite the increased variability in responses at intermediate ΔFs, consistent trends were clearly present in the average data. In particular, at all ΔFs but the smallest one, the probability of a "two streams" response increased noticeably over time. The increase was slower at the 3 than at the 6 and 9 semitone ΔFs. Also, the asymptotic probability of segregation was lower in the 3 than in the 6 and 9 semitone ΔF condition. On the whole, these results are consistent with those of earlier studies demonstrating a slower build-up, and intermediate probabilities of segregation, at intermediate than at large ΔFs, and no build-up at very small ΔFs (Anstis and Saida, 1985; Bregman, 1978; Carlyon et al., 2001).

Neural Responses to Tone Sequences

Alternating Tones

The responses of single units in the primary auditory cortex (A1) of two awake rhesus monkeys (*Macaca mulatta*) were recorded while the animals were listening to sound sequences similar to those used in the preceding psychophysical experiment in human listeners. The

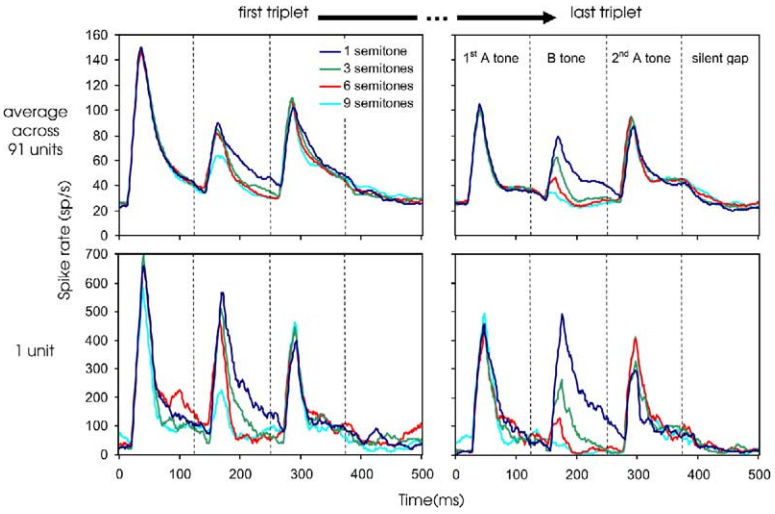

Figure 2. Poststimulus Time Histograms of Neural Responses to ABA Triplets

(Left) Neural responses to the first triplet in the sequence. (Right) Neural responses to the last (20th) triplet in the sequence. (Top) Average response across all 91 units recorded in the two animals. (Bottom) Example response from a single unit. The vertical dashed lines mark the onset or offset times of the individual tones, relative to the response. The temporal positions of the A and B tones and of the intertriplet gap are indicated in the upper right panel. The different colors indicate responses obtained using different frequency separations (ΔFs) between A and B, as shown in the legend in the upper left panel. The frequency-color coding scheme is the same as in Figure 1B. For illustration purposes, the PSTHs were smoothed using a 19 ms rectangular running-average window.

frequency of the A tone was adjusted to correspond to the best frequency (BF) of the unit being tested, as estimated from preceding tuning-curve measurements—it varied between 500 and 9514 Hz, with a geometric mean of 1663 Hz, and in over 80% of cases it was within 500 and 4000 Hz (the range of A tone frequencies used in the psychophysical experiment). As in the psychophysical experiment, the frequency of the B tone was set 1, 3, 6, or 9 semitones above that of the A tone and remained constant within a sequence.

Two major trends were apparent in the poststimulus time histograms (PSTHs) of neuronal responses to the ABA tone patterns (Figure 2). First, the responses to the B tones generally decreased as ΔF increased. This effect can be explained simply in terms of frequency selectivity: since the neurons, by definition, had their BF at the A tone frequency, increasing ΔF amounted to moving the B tones further and further away from the most sensitive region of the frequency tuning curve. A neural mechanism that may have further contributed to reducing the response to the B tone as ΔF increased is poststimulus adaptation or suppression following the first A tone from each triplet. Some evidence for this comes from the slight increase in the response to the second A tone as ΔF increased, which probably reflects a release from adaptation or from inhibition by the preceding B tone. This observation is consistent with earlier electrophysiological data on neural responses to tone sequences in AC (Bee and Klump, 2004; Brosch and Schreiner, 1994; Calford and Semple, 1995; Fishman et al., 2001, 2004; Kanwal et al., 2003).

The second major trend was a general decrease in response magnitude between the first and the last triplets (compare left and right panels in Figure 2). This response decay was general in the sense that it affected both A and B tone responses, at all frequency separations. In order to distinguish this effect from other forms of neural "adaptation," such as the adaptation of A tone responses by the preceding B tones described above, we will hereafter use the term "habituation" when referring to this longer-term decay in neural response, with no assumption regarding the underlying physiological mechanism at this stage.

To quantify this habituation and characterize its time course, we counted the number of spikes evoked in response to each of the successive tones in the 10 s stimulus sequence and studied how the counts varied as a function of time after sequence onset (Figure 3A). Overall, spike counts were found to decrease significantly as a function of triplet position (i.e., from the first to the last triplet in the sequence stimulus) [$F(1, 90) = 41.63$, $p < 0.0005$]. It is worth noting that although the largest decay in neural response generally occurred between the first and second triplets (at least for B tones), the responses often continued to decrease more slowly after that. The decay in spike counts was general, in the sense that it affected the responses to the A tones like those to the B tones, and was observed at all ΔFs.

Reversed Tone Sequences

We also measured the responses of roughly a third (i.e., 33) of the neural units to repeating triplets in which the temporal positions of the A and B tones were switched, resulting in BAB instead of ABA triplets. This was done in order to check that there was no major influence of the relative temporal positions of the A and B tones that could not be explained by BF and that the pattern of neural activity at cortical sites with BFs corresponding to the middle (formerly, B) tones was simply the converse of that observed at sites most responsive to the outer (formerly, A) tones. This was borne out in the data. The spike counts evoked by these reversed (BAB) triplet sequences, which are shown in Figure 3B, were essentially as expected based on those observed in Figure 3A, the responses to the B tones decreasing markedly with increasing ΔF, while the responses to the A tones (now in middle position) remained generally strong. From these data and those shown in Figure 3A, we can infer that when ΔF was small (e.g., 1 semitone), A1 units with BFs corresponding to the A or B frequency simultaneously produced large responses; in contrast, when ΔF was large (e.g., 6 or 9 semitones), two A1 populations were activated alternately by the A and B tones. Like those evoked by the original (ABA) sequences, the spike counts evoked by the B tones in the B-only sequences decreased significantly as a function of the triplet position within the sequence [F(1,

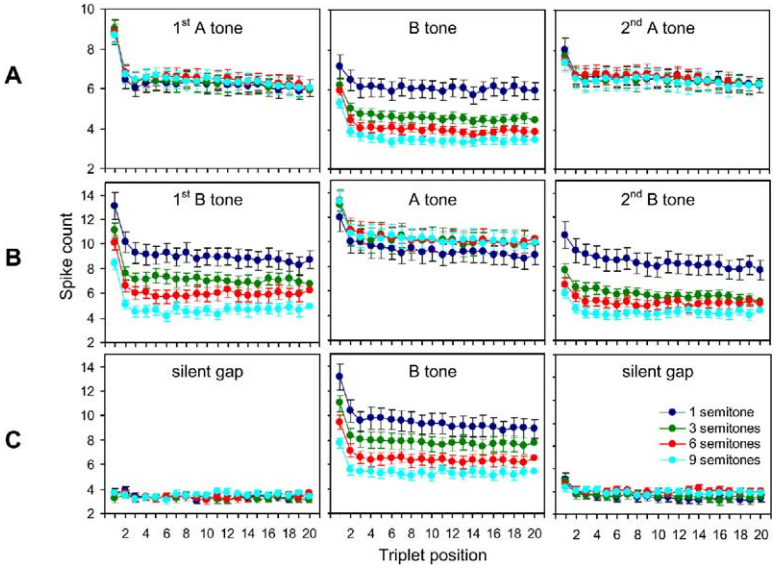

Figure 3. Spike Counts

(A) Spike counts evoked by the A and B tones in repeating ABA triplets as a function of the triplet position in the sequence. (Leftmost panel) Spike counts for the initial A tone (in each triplet). (Middle panel) Spike counts for the B tones. (Rightmost panel) Spike counts for the final A tone (in each triplet).

(B) Same type of results for inverted (i.e., BAB instead of ABA) triplets, so that the middle panel now corresponds to the middle A tone, while the left and right panels correspond to the two B tones that started and ended each triplet, respectively.

(C) Spike counts evoked by B tones in sequences containing only these tones—all the A tones being replaced by silent gaps of equivalent duration. The error bars in the different panels indicate 68% confidence intervals around the mean spike counts, as estimated using statistical bootstrap. The color coding scheme is the same as in previous figures.

32) = 49.75, $p < 0.0005$]; thus, in all cases, neural responses decayed over the course of the 10 s stimulus sequence.

Same-Frequency Tone Sequences

In order to determine whether the habituation of neural responses was critically dependent on the presence within the stimulus sequence of tones with different frequencies and/or on the use of a relatively high tone presentation rate, we also recorded neural responses to sequences consisting of just the B tones at the same temporal positions as in the ABA sequences—the alternate tones being replaced by silent gaps of equivalent durations. Complete data collection for this additional set of stimulus conditions could be achieved in 41 units. The corresponding mean spike counts are illustrated in Figure 3C. The spike counts evoked by the B tones in this condition were significantly larger than those evoked by the same tones in the full ABA-triplet condition [$F(1, 40) = 29.54$, $p < 0.0005$], an effect that can be explained by a release from forward inhibition by the A tones when these tones were replaced by silent gaps. The spike counts decreased significantly across triplet positions [$F(1, 40) = 81.78$ $p < 0.0005$], indicating that in this condition, like in previous ones, neural-response habituation occurred.

From Spikes to Percepts: A Simple Model of Auditory Streaming

In order to relate meaningfully the neural responses illustrated in Figures 2 and 3 to the psychophysical data in Figure 1B, we devised a simple model of how spike trains evoked by the ABA tone sequences could lead to the percept of one or two auditory streams. The problem of relating stochastic neural responses to probabilistic perceptual judgments (or behavioral responses) has been successfully addressed in earlier studies using the conceptual tools of statistical signal detection theory (Green and Swets, 1966; Johnson, 1980; Newsome et al., 1989; Parker and Newsome, 1998). The model proposed here is cast in this theoretical framework. The basic idea behind the model is that the neural activation patterns evoked in A1 by incoming sounds are "read out" by other neurons or groups of neurons, which behave as binary classifiers: based on the information they receive from A1, these neurons assume one of two possible states, which correspond to the "one stream" and "two streams" percepts (or responses). A key assumption in the model is that the decision between one and two streams is made by comparing the spike counts evoked by the A and B tones in each triplet to a fixed threshold; when the spike counts evoked by the B and both A tones exceed the threshold, a "one stream" response is produced; in contrast, when the A tones, but not the B tone, are detected, a "two streams" response is counted. This threshold is the only free parameter in the model, and although its value was adjusted so as to optimize the fit between data and predictions (using a maximum-likelihood criterion), it depended in no way on ΔF or time after sequence onset. Further details on the functioning of the model are given in the Experimental Procedures section.

The probabilities of "two streams" responses predicted as a function of both ΔF and time based on the measured neural responses, or "neurometric functions," are shown in Figure 4. The human psychometric functions from Figure 1B are replotted here as dashed lines to facilitate comparisons. It can be seen that the neurometric functions replicate all the major features of the psychometric functions measured in human listeners. Specifically, (1) at all ΔFs except the smallest, the predicted probability of a "two streams" response increased over time; (2) this increase was longer at the two intermediate ΔFs, 3 and 6 semitones; (3) at the largest ΔF (9 semitones), the predicted probability of a "two streams" response increased abruptly after sequence onset. This good qualitative agreement between the neurometric and the psychometric functions was paralleled by a good quantitative agreement, as demonstrated by the fact that the former usually fell

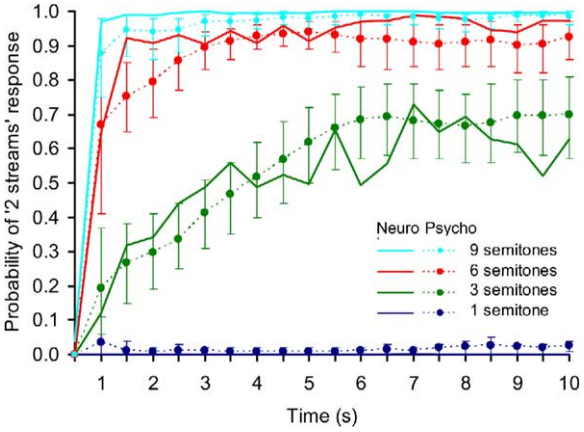

Figure 4. Comparison between Psychometric and Neurometric Functions Representing Probabilities of "Two-Stream" Responses as a Function of Time and ΔF

The psychometric functions from Figure 1 are replotted here as dashed lines, to facilitate comparison with the neurometric functions, which are shown as solid lines. The error bars indicate 95%-confidence intervals around the mean proportions estimated using statistical bootstrap.

within the 95% confidence intervals around the latter; thus, the predictions based on neural data from the two monkeys fall within the range of the across-subject variability in the psychophysical data.

Influence of Stimulus Presentation Rate

The previous discussion demonstrated that a model based on neural spike counts is consistent both with the effect of ΔF and of time on the streaming judgments made by human subjects. Here, we show that the same model can be used to explain another important psychophysical feature of auditory stream segregation, namely, its dependence on tone-presentation rate. The dashed lines in Figure 5 show psychometric functions measured using tone sequences similar to those used above, but in which the tone-presentation rate was decreased by inserting 50 ms silent gaps between the tones within each triplet and by increasing the intertriplet interval by the same amount. Comparing these functions to those shown in Figure 1B, it can be seen that, consistent with earlier findings (e.g., Bregman, 1990), this resulted in reducing the probability that listeners experience segregation and in a slower build-up of streaming, especially at the 3 semitone ΔF. The conditions that were the least affected by the change in tone presentation rate were the 9 semitone ΔF, where the average proportion of "two streams" responses remained high, and the 1 semitone ΔF, where the proportion of "two streams" responses was already very low prior to the reduction in tone-presentation rate.

In order to test whether these effects could be captured by a model similar to that described above, we used the data from the subset of 34 units for which responses were measured both to the (fast rate) ABA sequences and to sequences of B tones separated by 375 ms silent gaps. The response to the B tones in these two conditions can be thought of as representing two extreme values of intertone gap, in which the B tone was either immediately preceded by an A tone or occurred after a 375 ms silent interval. To estimate how the spike-count distributions varied as a function of tone-presentation rate, we independently fitted the spike-count distributions to the B tones in these two conditions with Gaussian probability density functions, resulting in two sets of best-fitting parameters. Linear interpolation between the two sets of parameters was then used to generate spike-count distributions for the slower ABA sequence used in the second psychophysical experiment; this condition has an intermediate tone-presentation rate, where each B tone was separated by the previous A tone by 50 ms. The position of the point at which the interpolation between the two sets of parameters was taken was the only free parameter used in the modified model, and its value was adjusted to yield the best possible (i.e., maximum likelihood) fit between the neurometric and psychometric functions in the faster-rate condition. The value of the spike-count threshold parameter was fixed, based on the results of the baseline (i.e., faster-rate) condition, and depended in no way on the neural or the psychophysical results of the slower-rate condition considered here. Finally, the interpolated spike-count distributions were fed to the same model as above.

The best-fitting neurometric functions in the slower-rate condition are shown as solid lines in Figure 5. It can be seen that the model predictions captured most of the psychophysical effects induced by the change in stimulation rate. In particular, it predicted larger decreases in the probability of "two streams" responses

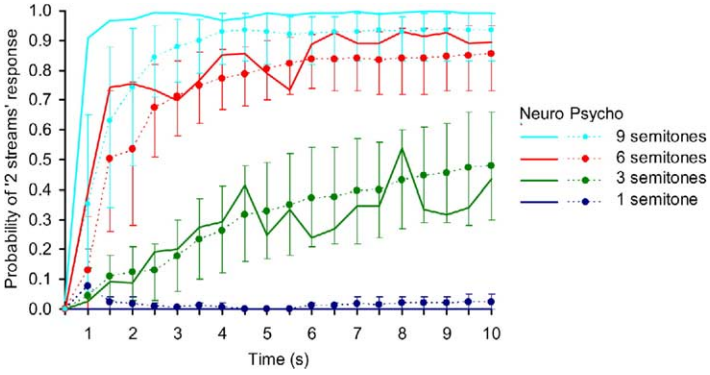

Figure 5. Comparison between Psychometric and Neurometric Functions for a Slower Tone Presentation Rate

The format is the same as that of Figure 4. The error bars indicate 95%-confidence intervals around the mean proportions estimated using statistical bootstrap.

and a slower build-up rate, at intermediate ΔFs (3 and 6 semitones). One aspect of the data, which the model did not adequately reproduce, is the slight reduction in build-up rate at the largest ΔF (9 semitones). However, on the whole, the model was rather successful in reproducing the psychometric functions for the slower-rate condition, as model predictions usually fell within the 95% confidence intervals around the observed probabilities of "two streams" responses.

Discussion

The present results demonstrate that neural responses recorded in A1 while awake primates listen to tone sequences can account both qualitatively and quantitatively for several essential features of an important perceptual auditory organization phenomenon in humans. The predictions derived from the neural data replicate successfully the increase in the probability of perceiving a tone sequence as two segregated auditory streams when the frequency separation (ΔF) between consecutive tones is increased. More importantly, the neurometric functions faithfully replicate the observed increases in the probability of hearing segregation (i.e., two streams) during the first 10 s following sequence onset at intermediate and large ΔFs. This result is noteworthy, because, unlike those obtained in previous studies (Bee and Klump, 2004; Fishman et al., 2001, 2004; Kanwal et al., 2003), it demonstrates a correspondence between changes in neural responses and changes in perception, which occur without any concomitant change in the physical stimulus. In addition, the same model that was used to relate the stochastic neural responses to the probabilistic perceptual judgments across two stimulus dimensions (ΔF and time since sequence onset) could also be applied to account for the observed influence of a third important stimulus parameter, i.e., tone-presentation rate. Altogether, these results demonstrate a remarkable correspondence between neural responses to tone sequences in A1 and perceptual judgments of stream segregation across several stimulus dimensions. Although this good correspondence does not prove that auditory streaming percepts are necessarily determined in A1 (see below), at the very least it shows that neural responses in A1 can account for several important features of the auditory streaming phenomenon.

Multisecond Habituation of A1 Responses and the Build-Up of Stream Segregation

The results obtained in this study suggest that a key mechanism behind the build-up of auditory streaming (Anstis and Saida, 1985; Bregman, 1978; Carlyon et al., 2001) is the "adaptation" or "habituation" of neural responses over time. In the model proposed here, the build-up of stream segregation is a by-product of that habituation. Neural-response habituation is an important phenomenon, likely to subserve many sensory-system functions (see for example, Fairhall et al., 2001). In vision, for instance, the phenomenon has been implicated in contrast gain control (Ohzawa et al., 1982) and information-transmission optimization (Muller et al., 1999). In the auditory modality, multisecond habituation in A1 has recently been implicated in the detection of deviant events inside temporal acoustic sequences and in the generation of the mismatch negativity (MMN) in auditory-evoked potentials (Ulanovsky et al., 2003, 2004). The present results suggest that this neural phenomenon also plays a role in the build-up of auditory stream segregation. Whether the build-up of stream segregation serves a functional role in itself remains uncertain. It has been proposed that the inertia in going from one percept to the other was required to maintain perceptual stability despite random fluctuations in neural activity (Bregman, 1990). On the other hand, that stream segregation takes time to occur may hamper adaptation in natural settings, where the rapid parsing of sounds into streams could be an important prerequisite for survival.

The finding that the characteristics (i.e., time course and extent) of neural habituation in A1 are compatible with those of the psychophysical build-up of auditory stream segregation has potential implications beyond the auditory modality. It has been suggested that the build-up of auditory stream segregation is the auditory analog of a visual-perception phenomenon known as the "breakdown of apparent motion" (Anstis et al., 1985): when two dots occupying different spatial positions are flashed in an alternating fashion, under appropriate stimulus conditions, viewers initially report perceiving a single dot moving back and forth (visual analog of a tone jumping up and down in pitch); after several seconds of continuous viewing, however, the percept switches to that of two stationary, flickering dots (analog of two parallel streams of constant-frequency tones). These analogous auditory "build-up" and visual "break-down" phenomena have traditionally been explained by the fatigue of specialized feature (e.g., motion) detectors under prolonged stimulation. According to this view, the build-up of stream segregation would be explained by the fact that (1) there are frequency-change detectors in the auditory system, (2) the activation of these detectors is needed for the perceptual integration of consecutive tones of different frequency, and (3) this activation decays during prolonged stimulation (Anstis and Saida, 1985; van Noorden, 1975). The present results suggest a simpler explanation, which does not require the habituation of, specifically, frequency-change detectors in the auditory system for the sensation of integrated pitch motion to be lost. Instead, stream segregation and its build-up appear to arise simply as a by-product of sound-event detection within frequency-specific, but otherwise unspecialized, neural populations in A1 and their habituation. It would be interesting to determine whether the same type of model, when applied to neural responses from primary visual cortex (V1) to sequences of alternating flashes, could also account for the breakdown of apparent motion in vision without necessarily having to invoke the fatigue of specialized motion detectors beyond V1.

Influence of Attention

The influence of attention on neural responses and perception is a traditionally important question in cognitive neuroscience. The animals in the present study performed a behavioral task, which required that they pay

at least some attention to the auditory stimuli during the electrophysiological recording sessions—just like the human listeners were required to pay attention to the stimuli during the psychophysical measures. An interesting question is whether the neural phenomena observed here, and the good agreement between the neural and psychophysical results, would have also been found if the animals had not been given this task. The only two previous reports of multisecond habituation of neural responses to tone sequences in A1 were in anesthetized animals (Ulanovsky et al., 2003, 2004). To the extent that the habituation observed in the present study is of a similar nature as that identified in those earlier reports, it should also manifest itself even when the animal is not actively attending to the auditory stimulus. This would be consistent with experimental results suggesting that stream segregation is largely an automatic, "primitive" perceptual organization phenomenon, which can occur even in the absence of attention (Macken et al., 2003; Näätänen et al., 2001; Sussman et al., 1999). On the other hand, it has been shown that if a listener's attention is momentarily directed away from the tone sequence after stream segregation has been allowed to fully build up, the percept is "reset" (i.e., goes back to that of "one stream") and the build-up starts all over again (Carlyon et al., 2001). Based on the present results, a simple explanation for this surprising effect is that when the listener's attention is directed toward the tone sequence, there is a momentary overall increase in neural responsiveness (Benson and Hienz, 1978), which causes the spike rates evoked by tones at and away from the BF to both exceed the detection threshold, resulting in a one stream percept; then, due to habituation, the responses start to decrease and the weaker spike rates evoked by tones away from the BF eventually fall below the detection threshold, while the higher spike rates evoked by BF tones remain above it, resulting in the percept of segregated streams.

The model proposed here can also account, in principle, for the influence of the listener's attentional set on auditory stream segregation. For instance, it has been shown that the ΔF above which listeners cannot help but hear the sequence as segregated even though they are trying to hold on to the percept of a single stream is larger than the ΔF below which listeners can only hear the sequence as a single stream despite their efforts to hear it as two (van Noorden, 1975). A simple explanation for these effects, based on the thresholding model proposed here, is that task instructions influence the listener's internal criterion (i.e., the threshold) for deciding between one and two streams: when the listener is actively trying to hear the sequence as a single stream, the threshold is lowered so that spike counts evoked by tones away from BF fall above it; conversely, when the listener is actively trying to hear two streams, the threshold is raised so that the spike counts evoked by tones away from the BF no longer exceed it. Obviously, for listeners as for the model, the range over which the threshold can vary is bounded downward by the fact that too low a threshold results in unrealistically high rates of false alarms and upward by the fact that too high a threshold leads to unacceptably high miss rates. This may explain why listeners have only limited control over the way in which they can hear sound sequences: beyond certain parameter values, they can only perceive the stimulus in one way (van Noorden, 1975).

The Brain Locus of Auditory Streaming

Although the results presented here demonstrate that neural responses in A1 can readily account for most of the important features of auditory streaming, including its build-up, this does not prove that A1 is necessarily the first, the final, or the only stage of processing involved in auditory streaming. The neural phenomena described here, including the habituation of responses over time, might be already present below the auditory cortex. Future studies will have to determine whether neural response characteristics (including the dependence of spike-count distributions on both ΔF and time after sequence onset) in the auditory nerve, cochlear nucleus, or the inferior colliculus are, like those measured in A1, qualitatively as well as quantitatively consistent with the measured characteristics of auditory streaming.

On the other hand, cortical areas other than A1 may also be involved in the perceptual organization of sound sequences into streams. In fact, the model proposed here posits that the perceptual decision between one and two streams is based on the activity of neurons or neural structures that "read out" the neural responses evoked in A1. These neurons could be located in auditory cortical areas beyond A1 or areas of the prefrontal cortex, which have been implicated in an auditory object-identification ("what") pathway that originates in the antero-lateral belt in nonhuman primates (Kaas and Hackett, 2000; Rauschecker and Tian, 2000; Rauschecker, 1998; Romanski et al., 1999; Tian et al., 2001) or the equivalent areas in humans (Binder et al., 2000, 2004; Rauschecker and Tian, 2000; Wessinger et al., 2001; Zatorre et al., 2004). A recent study using functional magnetic resonance imaging in humans also implicates the intraparietal sulcus in the perceptual formation of auditory streams (Cusack, 2005).

Limitations of the Present Study and Perspectives

One limitation, which the present work shares with previous studies of the intracortical basis of stream segregation, is that the animals' percepts were not measured. Therefore, the comparison between the neural and the psychophysical data rests on the assumption that the two species in which these data were obtained have similar auditory system and auditory perception characteristics. This assumption is supported by results in the literature, which show a close correspondence between human and primate auditory-cortex anatomy and function (Kaas and Hackett, 2000; Rauschecker and Tian, 2000; Wessinger et al., 2001) and similar auditory perceptual abilities (Pfingst, 1993), with some behavioral evidence that monkeys experience auditory stream segregation (Izumi, 2002) and other perceptual illusions (Miller et al., 2001; Petkov et al., 2003). Admittedly, psychometric functions for the build-up of stream segregation have never been measured in monkeys, and they might differ in their details from those measured in humans. However, to the extent that

the differences are not too pronounced, the model described here, and the main conclusions of the present study, should remain basically valid.

Recording the animals' percepts while they are listening to sound sequences might be achieved in future studies. However, getting animals to indicate their percept and to update their responses in "real time" as the sequence progresses will undoubtedly prove to be a challenging task for experimenters. From this point of view, the alternative approach adopted recently by some authors, which involves the conjoint measurement of psychophysical streaming judgments and brain activity in humans using functional brain-imaging techniques or magnetoencephalography (Cusack, 2005; Gutschalk et al., 2005), seems perhaps more promising. Unfortunately, the easier access to the listener's percepts in such human studies comes at the expense of limited insight into the underlying neural micromechanisms of perception.

Experimental Procedures

Psychophysical Experiment

Ten human subjects (5 male, 5 female, aged 21–42 years) with normal hearing (i.e., pure-tone thresholds below 20 dB HL [hearing loss] at octave frequencies between 250 and 8000 Hz) took part. The subjects were tested individually in a double-wall soundproof cabin (IAC). The stimuli were sequences of repeating ABA tone triplets, where A and B represent tone of different frequencies, as illustrated in Figure 1A. Each sequence was comprised of 20 triplets. Each tone was 125 ms in duration, including 20 ms raised-cosine ramps. In the basic test condition, there was no silent gap between the tones within a triplet, and the silent gap between consecutive triplets was also 125 ms long, resulting in a total sequence duration of 10 s. The A and B tone frequencies were kept constant within a sequence. The A tone frequency was selected randomly for each listener among four possible frequencies (500, 1000, 2000, and 4000 Hz). The B tone frequency was varied parametrically across sequences, and depending on the condition being tested, it was 1, 3, 6, or 9 semitones above the A tone frequency (for instance, at the 1000 Hz A tone frequency, the frequency of the B tones was 1059, 1189, 1414, or 1682 Hz). These four frequency separations (ΔFs) were selected, based on published data (Carlyon et al., 2001) and preliminary results, to yield different build-up rates and asymptotic levels of perceived stream segregation. Another series of test conditions was produced by inserting 50 ms silent gaps between the tones within a triplet and increasing the intertriplet gap duration to 175 ms, resulting in a slower tone-repetition rate. These additional stimulus conditions were tested in eight of the listeners. The stimuli were generated digitally under Matlab (The Mathworks, MA) and stored as 16 bit files on computer hard disk. They were played out at a 44.1 kHz sampling rate using a Lynx Studio Card, attenuated (TDT PA4, Tucker Davis Technology), preamplified (TDT HB6), and delivered diotically through headphones (Sennheiser HD250) at 70 dB SPL. The repeating tone-triplet sequences corresponding to the four different ΔFs were presented 20 times each, in a completely randomized order. Listeners were instructed to press "1" when they heard a single stream with a galloping rhythm and "2" when they heard two monotonic streams with different tempi. They were encouraged to start responding as early as possible after sequence onset and then to press the "1" or "2" key as soon and as often as they heard the percept change before the end of the 10 s sequence. The timing of key presses relative to sequence onset was measured and recorded. Each key press caused a switch in the value of a binary variable, from 0 ("one stream") to 1 ("two streams"), or vice versa. The binary values were sampled at 1 ms intervals over the entire sequence duration. Vectors corresponding to different presentations of the same sequence (20 presentations at each ΔF) were then averaged to obtain the probability of a "two streams" response as a function of time.

Electrophysiological Experiment

Two male rhesus monkeys aged between 6 and 8 years took part in electrophysiological recordings. They were implanted stereotactically with a recording cylinder over their left cerebral hemisphere. The position and orientation of the recording chamber were verified with magnetic resonance imaging (MRI), and the targeted recording areas inside the recording chamber were determined to include parts of the supratemporal plane, the lateral surface of the superior temporal gyrus, and the dorsal bank of the superior temporal sulcus. A plastic matrix grid (Crist Instruments) was fitted into the recording cylinder to facilitate electrode positioning. Lacquer-insulated tungsten microelectrodes (<10 MOhm) were inserted through the grid and into the cortex using a hydraulic microdrive (David Kopf Instruments). Action potentials were sent through a "slicer" unit, which selectively amplified all transients above a given amplitude, thereby enhancing the signal-to-noise ratio, if necessary. Subsequently, the signal was sent through a window discriminator, in order to permit reliable discrimination of spikes of different amplitude. The resulting trigger pulses were then transferred to a personal computer for data collection using the CORTEX program (NIMH, Bethesda, MD).

During electrophysiological recording, the animals were given a simple behavioral task, which they had been trained to perform prior to the experiment. The aim of this task was to ensure that the animals were awake and attentive to the auditory stimuli during the collection of the neural data—just like the human listeners were during the collection of the corresponding psychophysical data. The simple auditory discrimination task required that the animal touch with its hand a metal bar in front of it and release this bar within 1.0 s only when hearing a conditioned auditory stimulus (S+)—a short melody, consisting of four notes (CEGC with C at 512 Hz on a tempered scale), each of 100 ms duration. The conditioned stimulus was presented at random times in-between the streaming sequences, so that its occurrence was not predictable and the animal had to remain attentive throughout the testing session. Positive reinforcement (in the form of water delivered directly into the mouth through a tube) was conditioned upon release of the bar within a few seconds of hearing the conditioned stimulus. The animal was to hold the bar to all other sounds (S-), for which no water reward was given. If the animal failed to respond to the conditioned stimulus within 1 s, it would receive a 3–5 s time-out, after which stimulus presentation was restarted. If the animal failed to respond to the conditioned stimulus in several consecutive trials, the experimenter would intervene; usually, he entered the soundproof room, interacted with the animal, and decided when and whether to resume data collection. During recording and testing, the monkey's face and hand were monitored on a closed-circuit TV monitor, so that the experimenter had immediate control of the monkey's working condition. All recording sessions took place in the afternoon, and on the morning prior to a recording session the animals had only restricted access to water. (All procedures were approved by the Georgetown University Animal Care and Use Committee and conform to NIH standards.) Delivery of the conditioned stimulus and collection of behavioral responses were controlled by the CORTEX software (NIMH, Bethesda, MD). Behavioral responses were analyzed offline using Matlab (The Mathworks, MA) routines developed at NIMH and other laboratories.

Test stimuli consisted of isolated probe tones for neural frequency-response function measurement and various sequences of repeating, constant- or alternating-frequency tone sequences. The frequency-response functions of the units were measured using probe tones ranging in frequency between 500 and 20,000 Hz, in 24th octave steps. A bracketing procedure was used to determine the unit's best frequency (BF). The BFs of the units used in this study ranged between 500 and 9514 Hz, with a mean of 1663 Hz. Once a unit's BF was estimated, the unit's responses to 10 s sequences of ABA tone triplets, similar to those used in the psychophysical experiment in humans, were measured. These sequences consisted of 20 ABA tone triplets, with a 125 ms tone duration and a 125 ms silent gap between triplets. The frequency of the A tones was set to the unit's estimated BF. As in the psychophysical experiments, the frequency of the B tones was set 1, 3, 6, or 9 semitones above that of the A tones, and the resulting stimulus sequence was presented 20 times each, in randomized order. In 33 of the units,

responses to stimulus sequences in which the temporal positions of the A and B tones were switched (i.e., sequences of BAB triplets) were also recorded, in addition to the responses to the original ABA triplet sequences. Furthermore, in 34 units, responses to stimulus sequences consisting of just the A or just the B tones (with the alternate tones replaced by silent gaps of equivalent durations) were also collected. The stimuli were generated with Matlab (The Mathworks, MA) and presented with the SIGNAL program (Engineering Design, MA). They were delivered through loudspeakers (Polk Audio) in a large double-walled acoustic chamber (IAC 1205A) with sound-dampening acoustic material (Sonex) attached to the wall. Quasi-free-field conditions were ensured in the recording room for the sound field measured at the location of the monkey's head (Tian et al., 2001). The stimulus level, at the monkey's head was approximately 70 dB SPL.

After recordings were completed, grand-average poststimulus time histograms (PSTHs) were computed by averaging neural responses across all 20 triplet positions within each sequence, as well as across all 20 sequence presentations at each ΔF, and further selection of units for analysis was performed. Complete recordings to ABA sequences were obtained in over 100 units overall in the two monkeys. Recording positions were assigned to cortical fields according to BF gradients (high-to-low from caudal to rostral in A1, see Rauschecker et al., 1995) and according to frequency/bandwidth tuning, response quality, and latency to pure-tone stimuli (core versus belt, see Rauschecker and Tian, 2004). Units whose estimated location was outside of A1 were eliminated from the analyses presented here. As a result, the data shown in Figures 2–4 are based on responses to ABA sequences from 91 units overall (28 in the first animal, 63 in the second), and on responses to A- or B-only sequences from 34 units (11 in the first animal, 23 in the second).

Subsequent analysis of the neural responses involved counting the number of spikes occurring during the presentation of each tone. The distributions of spike counts observed across the 20 presentations of each stimulus sequence (one sequence per ΔF) were then transformed into probability distributions (by dividing the number of times each possible spike count was observed by the total number of spike counts observed, so that the area under the distribution was equal to 1). This was done separately for each neuron, each tone, and each triplet position. Then, for each tone and each triplet position, the probability distributions corresponding to the different units were convolved with each other. This was done in order to determine the probability distribution of a random variable representing the total spike-count information obtained by performing an (unweighted) average of the spike counts from all available units, assuming statistical independence between units. The resulting spike-count probability distribution was used in order to derive the neurometric functions. The probability that a particular tone at a particular triplet position was detected was computed as the area under the probability distribution to the right of a predefined count, or "threshold." The resulting A and/or B tone-detection probabilities were subsequently used to estimate the probabilities of "one stream" and "two streams" responses, as follows. When both the B tone and the two A tones from the same triplet were detected by the model, a "one stream" response was counted. In contrast, if the two A tones were detected but the intervening B tone was not, a "two-streams" response was counted. This can be understood by considering that the decisions were based on the responses of units whose BF corresponded to A. If the presence of the B tone was detected in the responses of these units, then reciprocally, the presence of the A tones was likely to be detected in the responses of units whose BF corresponded to B (since the choice of setting the frequency of the A tones, rather than that of the B tones, to the unit's BF was arbitrary). Thus, in that situation, the two tonotopic sites corresponding to the A and B frequencies were not differentially activated, and the activity at both sites indicated (to units higher up) the presence of three consecutive events. This response pattern was interpreted as a single stream with a galloping rhythm. In contrast, if the presence of the B tone was not detected based on the responses of units whose BF corresponded to A, then it was likely by symmetry that the presence of the A tones was not detected based on the responses of units whose BF corresponded to B. In that situation, the two tonotopic sites corresponding to A and B in the AC were differentially activated, with activity at one site reflecting the presence of only the A tones and activity at the other site reflecting only the presence of the (half as numerous) B tones. This response pattern was interpreted as two streams with different tempi. Finally, the probabilities of "one stream" and "two streams" judgments (or responses) were evaluated simply by counting the number of corresponding decisions made by the model.

Best fits between the neurometric functions (representing the probability of a "two streams" response as a function of time) and the corresponding psychometric functions were performed by varying a single parameter: the threshold count above which a tone was declared detected. The threshold was not allowed to vary across ΔF or time after sequence onset.

Acknowledgments

Supported by Engineering and Physical Sciences Research Council Research grant GR/M90146 via HearNet (C.J. Darwin, Experimental Psychology, University of Sussex, UK), NIH R01 DC003489 (J.P.R., Georgetown University), NIH R01 DC05216 (A.J. Oxenham, RLE, MIT), and CNRS (C.M., France). The authors are grateful to J. Bernstein, B. Delgutte, L. Demany, J. McDermott, C. Lorenzi, J. Melcher, A. Norena, A. Oxenham, S. Shamma, C. Sumner, and three anonymous reviewers for helpful comments on an earlier version of this manuscript.

Received: March 10, 2005
Revised: July 15, 2005
Accepted: August 24, 2005
Published: October 5, 2005

References

Alain, C., Cortese, F., and Picton, T.W. (1998). Event-related brain activity associated with auditory pattern processing. Neuroreport *15*, 3537–3541.

Anstis, S., and Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. J. Exp. Psychol. *11*, 257–271.

Anstis, S.M., Giaschi, D., and Cogan, A.I. (1985). Adaptation to apparent motion. Vision Res. *25*, 1051–1062.

Beauvois, M.W., and Meddis, R. (1991). A computer model of auditory stream segregation. Q. J. Exp. Psychol. *43*, 517–541.

Bee, M.A., and Klump, G.M. (2004). Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. J. Neurophysiol. *92*, 1088–1104. Benson, D.A., Hienz, R.D. (1978). Single-unit activity in the auditory cortex of monkeys selectively attending left vs. right ear stimuli. Brain Res. *159*, 307–320.

Benson, D.A., and Hienz, R.D. (1978). Single-unit activity in the auditory cortex of monkeys selectively attending left vs. right ear stimuli. Brain Res. *159*, 307–320.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., and Possing, E.T. (2000). Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex *10*, 512–528.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., and Ward, B.D. (2004). Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. *7*, 295–301.

Bregman, A.S. (1978). Auditory streaming is cumulative. J. Exp. Psychol. Hum. Percept. Perform. *4*, 380–387.

Bregman, A.S. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound (Cambridge, MA: MIT Press).

Bregman, A.S., and Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. J. Exp. Psychol. *89*, 244–249.

Brosch, M., and Schreiner, C.E. (1994). Time course of forward masking tuning curves in cat primary auditory cortex. J. Neurophysiol. *77*, 923–943.

Calford, M.B., and Semple, M.N. (1995). Monaural inhibition in cat auditory cortex. J. Neurophysiol. *73*, 1876–1891.

Carlyon, R.P. (2004). How the brain separates sounds. Trends Cogn. Sci. *8*, 465–471.

Carlyon, R.P., Cusack, R., Foxton, J.M., and Robertson, I.H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. J. Exp. Psychol. Hum. Percept. Perform. *27*, 115–127.

Cusack, R. (2005). The intraparietal sulcus and perceptual auditory organization. J. Cogn. Neurosci. *17*, 641–651.

Fairhall, A.L., Lewen, G.D., Bialek, W., and de Ruyter Van Steveninck, R.R. (2001). Efficiency and ambiguity in an adaptive neural code. Nature *412*, 787–792.

Fay, R.R. (1998). Auditory stream segregation in goldfish (Carassius auratus). Hear. Res. *120*, 69–76.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., and Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear. Res. *151*, 167–187.

Fishman, Y.I., Arezzo, J.C., and Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J. Acoust. Soc. Am. *116*, 1656–1670.

Green, D.M., and Swets, J.A. (1966). Signal Detection Theory and Psychophysics (New York: Wiley).

Gutschalk, A., Micheyl, C., Melcher, J.R., Rupp, A., Scherg, M., and Oxenham, A.J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. J. Neurosci. *25*, 5382–5388.

Hartmann, W.M., and Johnson, D. (1991). Stream segregation and peripheral channeling. Music Percept. *9*, 155–184.

Hulse, S.H., MacDougall-Shackleton, S.A., and Wisniewski, A.B. (1997). Auditory scene analysis by songbirds: Stream segregation of birdsong by European starlings (Sturnus vulgaris). J. Comp. Psychol. *111*, 3–13.

Hung, J., Jones, S.J., and Vaz Pato, M. (2000). Scalp potentials to pitch change in rapid tone sequences. A correlate of sequential stream segregation. Exp. Brain Res. *140*, 56–65.

Izumi, A. (2002). Auditory stream segregation in Japanese monkeys. Cognition *82*, 113–122.

Johnson, K.O. (1980). Sensory discrimination: decision process. J. Neurophysiol. *43*, 1771–1792.

Jones, S.J., Longe, O., and Vaz Pato, M. (1998). Auditory evoked potentials to abrupt pitch and timbre change of complex tones: electrophysiological evidence of 'streaming'? Electroencephalogr. Clin. Neurophysiol. *108*, 131–142.

Kaas, J.H., and Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. Proc. Natl. Acad. Sci. USA *97*, 11793–11799.

Kanwal, J.S., Medvedev, A.V., and Micheyl, C. (2003). Neurodynamics of auditory stream segregation: tracking sounds in the mustached bat's natural environment. Network *14*, 413–435.

Logothetis, N., and Schall, D. (1989). Neuronal corelates of subjective visual perception. Science *245*, 761–763.

MacDougall-Shackleton, S.A., Hulse, S.H., Gentner, T.Q., and White, W. (1998). Auditory scene analysis by European starlings (Sturnus Vulgaris): perceptual segregation of tone sequences. J. Acoust. Soc. Am. *103*, 3581–3587.

Macken, W.J., Tremblay, S., Houghton, R.J., Nicholls, A.P., and Jones, D.M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. J. Exp. Psychol. Hum. Percept. Perform. *29*, 43–51.

McCabe, S.L., and Denham, M.J. (1997). A model of auditory streaming. J. Acoust. Soc. Am. *101*, 1611–1621.

Micheyl, C., Carlyon, R.P., Cusack, R., and Moore, B.C.J. (2005). Performance measures of auditory organization. In Auditory Signal Processing: Physiology, Psychoacoustics, and Models, D. Pressnitzer, A.d. Cheveigné, S. McAdams and L. Collet, eds. (New York: Springer), pp. 203–211.

Miller, C.T., Dibble, E., and Hauser, M.D. (2001). Amodal completion of acoustic signals by a nonhuman primate. Nat. Neurosci. *4*, 783–784.

Muller, J.R., Metha, A.B., Krauskopf, J., and Lennie, P. (1999). Rapid adaptation in visual cortex to the structure of images. Science *285*, 1405–1408.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. Trends Neurosci. *24*, 283–288.

Newsome, W.T., Britten, K.H., and Movshon, J.A. (1989). Neuronal correlates of a perceptual decision. Nature *341*, 52–54.

Parker, A.J., and Newsome, W.T. (1998). Sense and the single neuron: probing the physiology of perception. Annu. Rev. Neurosci. *21*, 227–277.

Petkov, C.I., O'Connor, K.N., and Sutter, M.L. (2003). Illusory sound perception in macaque monkeys. J. Neurosci. *23*, 9155–9161.

Pfingst, B.E. (1993). Comparison of spectral and nonspectral frequency difference limens for human and nonhuman primates. J. Acoust. Soc. Am. *93*, 2124–2129.

Ohzawa, I., Sclar, G., and Freeman, R.D. (1982). Contrast gain control in the cat visual cortex. Nature *298*, 266–268.

Rauschecker, J.P. (1998). Cortical processing of complex sounds. Curr. Opin. Neurobiol. *8*, 516–521.

Rauschecker, J.P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. USA *97*, 11800–11806.

Rauschecker, J.P., and Tian, B. (2004). Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. J. Neurophysiol. *91*, 2578–2589.

Rauschecker, J.P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. Science *268*, 111–114.

Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., and Rauschecker, J.P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci. *2*, 1131–1136.

Sussman, E., Ritter, W., and Vaughan, H.G., Jr. (1999). An investigation of the auditory streaming effect using event-related brain potentials. Psychophysiology *36*, 22–34.

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J.P. (2001). Functional specialization in rhesus monkey auditory cortex. Science *292*, 290–293.

Ulanovsky, N., Las, L., and Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. Nat. Neurosci. *6*, 391–398.

Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. J. Neurosci. *24*, 10440–10453.

van Noorden, L.P.A.S. (1975). Temporal Coherence in the Perception of Tone Sequences (Eindhoven: Eindhoven University of Technology).

Wessinger, C.M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., and Rauschecker, J.P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. J. Cogn. Neurosci. *13*, 1–7.

Zatorre, R.J., Bouffard, M., and Belin, P. (2004). Sensitivity to auditory object features in human temporal neocortex. J. Neurosci. *24*, 3637–3642.